HƯỚNG TỚI MÃ HÓA HÌNH ẢNH CHÍNH XÁC: CẢI THIỆN TẠO ẢNH TỰ HỒI QUY VỚI LƯỢNG TỬ HÓA VECTOR ĐỘNG

Ngô Trần Tuấn Anh[1]

[1] Trường Đại học Công nghệ Thông tin

Giới thiệu

Chúng tôi giới thiệu khung làm việc sinh ảnh với định lượng vector động (Dynamic VQ), cụ thể:

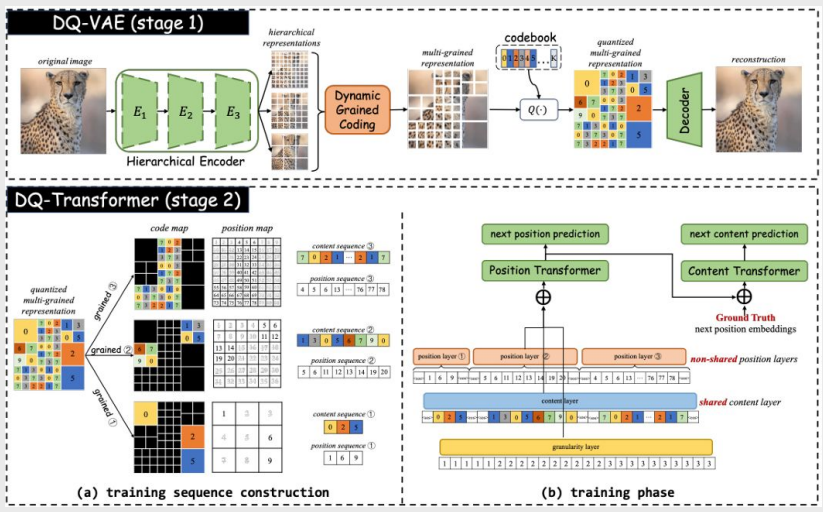
- **Đề xuất DQ-VAE:** Mã hóa độ dài biến thiên dựa trên mật độ thông tin vùng ảnh.
- **Phát triển DQ-Transformer:** Sinh ảnh tự hồi quy theo lộ trình từ thô đến mịn.
- **Kiểm chứng:** Tối ưu chất lượng (FID) và tốc độ vượt trội trên FFHQ và ImageNet.

Mục tiêu

Nhằm giải quyết vấn đề lãng phí tài nguyên và thiếu hụt chi tiết của mã hóa lưới cố định:

- **Tối ưu biểu diễn:** Phân bổ mã linh hoạt để tái tạo chi tiết sắc nét và giảm thiểu mã dư thừa.
- **Nhất quán cấu trúc:** Thay thế quét dòng bằng trình tự thô-mịn để đảm bảo bố cục ảnh đồng nhất.
- **Vượt mốc SOTA:** Nâng cao chất lượng hình ảnh sinh ra đồng thời tăng tốc độ suy luận thực tế.

Tổng quan mô hình



Hình 1: Sơ đồ tổng quát khung làm việc hai giai đoạn (DQ-VAE & DQ-Transformer)

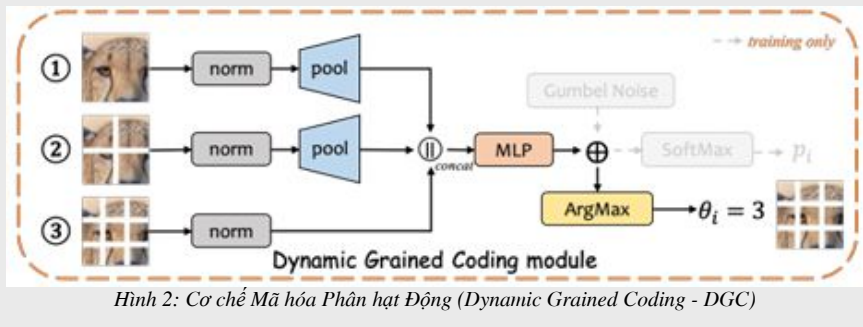
Khung làm việc gồm hai giai đoạn phối hợp để tối ưu hóa biểu diễn và sinh ảnh (*Hình 1*):

- **Giai đoạn 1 (DQ-VAE):** Sử dụng mô-đun DGC để mã hóa ảnh linh hoạt theo mật độ thông tin (vùng chi tiết được gán nhiều mã hơn vùng đơn giản).
- **Giai đoạn 2 (DQ-Transformer):** Sinh ảnh tự hồi quy từ Thô đến Mịn, dự đoán luân phiên Vị trí và Nội dung mã thông báo.
- **Luồng xử lý:** Ảnh gốc → Mã hóa phân hạt động → Dự đoán phân cấp → Ảnh tổng hợp sắc nét.

Phương pháp

BƯỚC 1: Mã hóa động với DQ-VAE

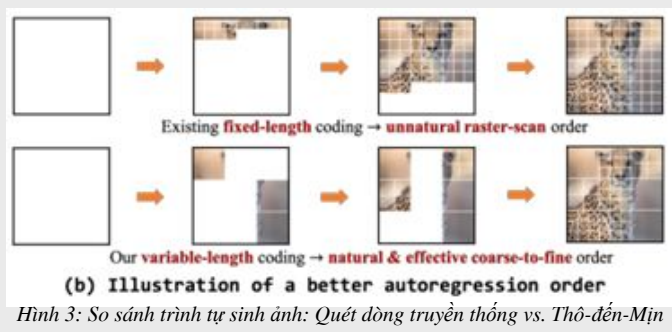
- **Cơ chế DGC:** Sử dụng mô-đun Dynamic Grained Coding để tự động gán độ dài mã biến thiên theo mật độ thông tin từng vùng ảnh. *(Hình 2)*
- **Tối ưu hóa:** Kết hợp mạng cổng (Gating) và Budget Loss để ưu tiên mã cho vùng chi tiết, đồng thời giảm dư thừa tại vùng mịn.
- **Kết quả:** Biểu diễn hình ảnh chính xác nhưng vẫn đảm bảo tính nhỏ gọn tối ưu.



Hình 2: Cơ chế Mã hóa Phân hạt Động (Dynamic Grained Coding - DGC)

BƯỚC 2: Sinh ảnh phân cấp DQ-Transformer

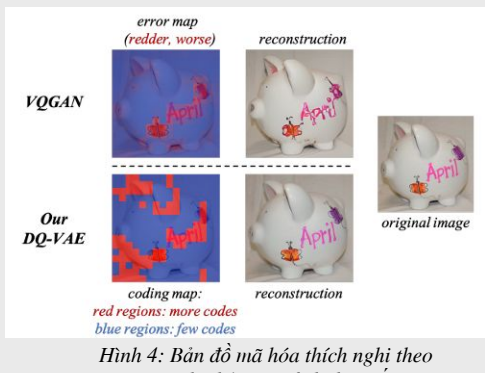
- **Stacked Transformer:** Dự đoán luân phiên giữa Vị trí (Position) và Nội dung (Content) của mã thông báo thông qua cấu trúc Transformer xếp chồng.
- **Trình tự Coarse-to-fine:** Sinh ảnh từ thô đến mịn, xây dựng khung sườn tổng thể trước khi lấp đầy các chi tiết cục bộ phức tạp. *(Hình 3)*
- **Cơ chế nhúng:** Sử dụng lớp nhúng Shared-content giúp mô hình hóa sự liên kết chặt chẽ giữa các cấp độ hạt khác nhau.



Hình 3: So sánh trình tự sinh ảnh: Quét dòng truyền thống vs. Thô-đến-Mịn (Coarse-to-fine)

BƯỚC 3: Tối ưu lộ trình và Hiệu năng *(Hình 4)*

- **Thay thế Raster-scan:** Loại bỏ quét dòng truyền thống, thay bằng lộ trình dựa trên nội dung giúp tập trung tính toán vào vùng quan trọng.
- **Tăng tốc suy luận:** Rút ngắn chuỗi mã giúp tăng tốc độ sinh ảnh từ 1.5x – 2x so với mô hình ViT-VQGAN truyền thống.
- **Nhất quán cấu trúc:** Lộ trình thô-mịn đảm bảo sự đồng nhất kết cấu, loại bỏ lỗi đứt gãy hình ảnh ở các mô hình tự hồi quy cũ.



Hình 4: Bản đồ mã hóa thích nghi theo mật độ thông tin ảnh thực tế

Ngô Trần Tuấn Anh – Trường Đại học Công nghệ Thông tin
Email : anhntt.20@grad.uit.edu.vn